

# PEMODELAN RISIKO PENYAKIT PNEUMONIA PADA BALITA DI PROVINSI JAWA TIMUR DENGAN PENDEKATAN *GEOGRAPHICALLY WEIGHTED LOGISTIC REGRESSION*

Evi Noviyantari Fatimah<sup>§1</sup>, I Komang Gde Sukarsa<sup>2</sup>, Made Susilawati<sup>3</sup>

<sup>1</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: evi.noviyantari@gmail.com]

<sup>2</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: sukarsakomang@yahoo.com]

<sup>3</sup>Jurusan Matematika, Fakultas MIPA - Universitas Udayana [Email: susilawati.made@gmail.com]

<sup>§</sup>*Corresponding Author*

## ABSTRACT

*This research is aim to determine the comparison of logistic regression models and models Geographically Weighted Logistic Regression and the factors that significantly affect the risk of pneumonia in toddlers in East Java Province. Logistic regression is a statistical analysis that is used to describe the response variable is categorical with the independent variables are categorical or continuous. The main problem of this method if it's applied in data that is affected of geographic location or spatial data. One of many method to solve the spatial data is Geographically Weighted Logistic Regression (GWLR). GWLR is a statistical method for analyze the data to account for spatial factor. The results showed that there are no significant differences between the logistic regression model with GWLR model. Factors that significantly affect the risk of pneumonia in toddlers in East Java Province is the percentage of low birth weight, the percentage of toddlers who get measles immunization, the percentage of toddlers who get vitamin A, and the percentage of toddlers who get DPT+HB immunization.*

*Keywords: Logistic Regression, Geographically Weighted Logistic Regression, Pneumonia*

## 1. PENDAHULUAN

Penyakit pneumonia merupakan salah satu penyebab dari kesakitan dan kematian pada anak di seluruh dunia. Pneumonia terjadi karena infeksi akut yang mengenai jaringan paru-paru, dengan gejala berupa batuk pilek yang disertai nafas sesak dan nafas cepat karena paru-paru yang meradang secara mendadak. Beberapa faktor risiko kesakitan dan kematian akibat penyakit pneumonia, yaitu faktor instrinsik dan faktor ekstrinsik (Noviana, *et al.* [5]).

Faktor instrinsik meliputi status gizi, berat badan lahir rendah, status imunisasi, pemberian ASI, dan pemberian vitamin A. Sedangkan Faktor ekstrinsik meliputi kepadatan tempat tinggal, polusi udara, tipe rumah, letak dapur, jenis bahan bakar, penggunaan obat nyamuk, asap rokok, penghasilan keluarga, jumlah

penduduk, serta faktor ibu baik pendidikan, umur, maupun pengetahuan ibu. Berdasarkan uraian di atas, maka sangat diperlukan upaya serius dalam menangani masalah pneumonia tersebut.

Salah satu cara yang dapat dilakukan untuk menurunkan angka kesakitan dan kematian pada anak adalah mengetahui faktor-faktor penyebabnya. Faktor-faktor penyebab terjadinya pneumonia pada tiap wilayah diduga berbeda-beda tergantung pada karakteristik dari masing-masing kabupaten atau kota tersebut. Oleh karena itu, diperlukan suatu metode pemodelan statistik dengan memperhitungkan faktor spasial untuk setiap pengamatannya. Metode statistika yang telah dikembangkan untuk menganalisis data dengan memperhitungkan faktor spasial adalah

### *Geographically Weighted Logistic Regression (GWLR).*

GWLR merupakan bentuk lokal dari regresi logistik dengan memperhatikan lokasi yang berasumsi bahwa data berdistribusi Bernoulli. Pendugaan parameter model GWLR yang diperoleh menggunakan *Maximum Likelihood Estimation* yaitu dengan memberikan suatu pembobot yang berbeda pada setiap lokasi.

Penelitian tentang GWLR pernah dilakukan oleh Anggarini & Purhadi [2] dan Pradita [6]. Dari penelitian tersebut dapat disimpulkan bahwa model GWLR dengan pembobot *adaptive bisquare kernel* dan pembobot *bisquare kernel* merupakan model terbaik untuk mengatasi adanya permasalahan faktor spasial atau lokasi geografis, karena memiliki nilai *Akaike Information Criterion* (AIC) yang lebih kecil dibandingkan model regresi logistik, dan penelitian tentang penyakit pneumonia pernah dilakukan oleh Santoso, *et al.*[7] yaitu tentang faktor-faktor eksternal pneumonia pada balita di Jawa Timur dengan pendekatan *Geographically Weighted Regression* (GWR). GWR merupakan suatu model regresi sederhana yang memperhitungkan faktor lokasi pengamatan dan menghasilkan penduga parameter model yang bersifat lokal untuk setiap titik dimana data tersebut diamati (Fotheringham, *et al.* [3]).

Berdasarkan uraian di atas, maka akan dilakukan analisis dengan menggunakan pendekatan GWLR untuk mendapatkan perbandingan model regresi logistik dengan model GWLR dan mengetahui faktor-faktor yang mempengaruhi risiko penyakit pneumonia pada balita tiap kabupaten/kota di Provinsi Jawa Timur tahun 2012.

Regresi logistik merupakan salah satu metode statistik yang dapat digunakan untuk menghubungkan variabel respon yang bersifat dikotomis (berskala nominal atau ordinal dengan dua kategori) atau polikotomis (berskala nominal atau ordinal dengan lebih dari dua kategori) dengan satu atau lebih variabel bebas yang bersifat kontinu atau kategorik (Agresti [1]). Pada regresi logistik,

untuk variabel responnya dianggap hanya mempunyai dua nilai yang mungkin, yaitu 0 (gagal) atau 1 (sukses), sehingga variabel respon ( $y$ ) tersebut mengikuti distribusi Bernoulli dengan fungsi peluang distribusi sebagai berikut:

$$P(Y = y) = \pi^y(1 - \pi)^{1-y}; \quad y = 0,1$$

dimana jika  $y = 0$  maka  $P(Y = 0) = 1 - \pi$  dan jika  $y = 1$  maka  $P(Y = 1) = \pi$ .

Bentuk umum dari model regresi logistik dapat dinyatakan sebagai berikut:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

Untuk mempermudah dalam pendugaan parameter regresi, maka  $\pi(x)$  ditransformasi dengan menggunakan transformasi logit sehingga diperoleh bentuk sebagai berikut (Fotheringham, *et al.* [3]):

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

Pengujian regresi logistik dilakukan dengan dua cara yaitu pengujian parameter secara serentak dilakukan untuk mengetahui signifikansi parameter  $\beta$  terhadap variabel respon dengan menggunakan statistik uji G dan pengujian parameter secara parsial dilakukan untuk mengetahui signifikansi parameter  $\beta$  terhadap variabel respon secara parsial dengan menggunakan statistik uji *Wald* (Hosmer dan Lemeshow [4]).

Model *Geographically Weighted Logistic Regression* (GWLR) merupakan model regresi yang memperhitungkan faktor spasial. GWLR adalah perkembangan dari metode GWR dengan variabel respon yang bersifat kategorik melalui regresi logistik. Pada model GWLR, variabel respon  $y$  diprediksi dengan variabel bebas yang masing-masing koefisien regresinya bergantung pada lokasi dimana data tersebut diamati (Fotheringham, *et al.* [3]). Secara umum bentuk dari model GWLR dapat dinyatakan sebagai berikut:

$$\pi(x_i) = \frac{\exp\left(\sum_{k=0}^p \beta_k(u_i, v_i)x_{ik}\right)}{1 + \exp\left(\sum_{k=0}^p \beta_k(u_i, v_i)x_{ik}\right)}$$

$$g(x_i) = \ln\left[\frac{\pi(x_i)}{1 - \pi(x_i)}\right]$$

$$= \beta_0(u_i, v_i) + \beta_1(u_i, v_i)x_{i1} + \dots + \beta_k(u_i, v_i)x_{ik}$$

Pengujian yang pertama dilakukan adalah pengujian kesamaan model regresi logistik dengan model GWLR untuk mengetahui apakah model GWLR lebih sesuai digunakan dibandingkan dengan model regresi logistik, dengan hipotesis sebagai berikut:

$$H_0 : \beta_k(u_i, v_i) = \beta_k ; k = 1, 2, \dots, k$$

$$H_1 : \text{minimal ada satu } \beta_k(u_i, v_i) \text{ yang}$$

berhubungan dengan lokasi  $(u_i, v_i)$

Statistik ujiannya:

$$F_{hitung} = \frac{\text{devians model 1/df}_1}{\text{devians model 2/df}_2}$$

Pengujian secara parsial dilakukan untuk mengetahui variabel mana saja yang berpengaruh signifikan terhadap variabel responnya dengan hipotesis yang diuji sebagai berikut:

$$H_0 : \beta_k(u_i, v_i) = 0$$

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

Statistik ujiannya:

$$z = \frac{\hat{\beta}_k(u_i, v_i)}{se(\hat{\beta}_k(u_i, v_i))}$$

Pemilihan pembobot spasial yang digunakan dalam menduga parameter sangat penting untuk menentukan besarnya pembobot masing-masing lokasi yang berbeda. Pembobot yang digunakan adalah fungsi kernel sebagai berikut (Fotheringham, *et al.* [3]):

a. Fungsi *Gaussian Kernel*

$$w_i(u_i, v_i) = \exp\left(-\frac{1}{2}(d_{ij}/h)^2\right)$$

b. Fungsi *Bisquare Kernel*

$$w_i(u_i, v_i) = \begin{cases} \left(1 - (d_{ij}/h)^2\right)^2, & \text{jika } d_{ij} < h \\ 0 & \text{jika } d_{ij} > h \end{cases}$$

dengan  $d_{ij}$  adalah jarak *Euclidean* antar lokasi  $i$

dan  $j$ , dengan persamaan:

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}, \quad h \text{ adalah}$$

parameter penghalus (*bandwidth*).

Untuk mendapatkan *bandwidth* optimum maka dapat dilakukan dengan menghitung *cross validation* (CV). Jika nilai CV semakin kecil maka akan didapatkan *bandwidth* yang optimum.

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2$$

## 2. METODE PENELITIAN

Data yang digunakan pada penelitian ini adalah data sekunder mengenai risiko penyakit pneumonia pada balita tahun 2012 yang diperoleh di <http://www.dinkes.jatimprov.go.id>, dengan unit pengamatan di 38 tingkatan kabupaten atau kota di Provinsi Jawa Timur.

Adapun variabel yang digunakan dalam penelitian ini sebagai berikut:

1. Variabel respon (Y) yaitu persentase balita penderita pneumonia di Provinsi Jawa Timur
2. Variabel bebas (X) yaitu persentase balita gizi buruk ( $X_1$ ), persentase berat bayi lahir rendah ( $X_2$ ), persentase balita yang mendapatkan imunisasi campak ( $X_3$ ), persentase balita yang mendapatkan vitamin A ( $X_4$ ), persentase pemberian ASI eksklusif pada balita ( $X_5$ ), dan persentase balita yang mendapatkan imunisasi DPT+HB ( $X_6$ )

Langkah-langkah yang dilakukan dalam metode penelitian ini adalah:

1. Melakukan analisa statistika deskriptif sebagai gambaran awal untuk mengetahui data tentang Risiko Penyakit Pneumonia pada Balita di Provinsi Jawa Timur
2. Memeriksa adanya multikolinieritas antara variabel-variabel bebas
3. Mendapatkan model regresi logistik dengan melakukan pengujian parameter secara serentak dan parsial

4. Menganalisis model GWLR dengan langkah sebagai berikut:
  - a. Menghitung jarak *Euclidean* antar lokasi pengamatan berdasarkan posisi geografis, menentukan *bandwidth* optimum dan menghitung matriks pembobot dari setiap lokasi pengamatan dengan menggunakan fungsi pembobot *gaussian* kernel
  - b. Menduga parameter model GWLR
  - c. Melakukan pengujian kesamaan model regresi logistik dan model GWLR
  - d. Melakukan pengujian parameter model GWLR dengan fungsi pembobot *gaussian* kernel
5. Membandingkan model regresi logistik dengan model GWLR

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Gambaran Pneumonia di Jawa Timur

Gambaran umum mengenai risiko penyakit pneumonia pada balita di Jawa Timur dapat dilihat pada Tabel berikut.

Tabel 1. Statistika Deskriptif Risiko Penyakit Pneumonia

Variabel	Mean	StDEv	Min	Max
Y	29,01	23,58	3,71	90,17
X <sub>1</sub>	1,199	0,741	0,210	3,700
X <sub>2</sub>	3,630	1,515	1,250	9,660
X <sub>3</sub>	99,39	6,99	78,67	112,27
X <sub>4</sub>	76,95	12,59	34,09	98,87
X <sub>5</sub>	64,20	10,68	37,82	87,08
X <sub>6</sub>	102,23	8,21	79,86	128,24

Tabel 1 menunjukkan bahwa persentase rata-rata balita penderita pneumonia (Y) di Provinsi Jawa Timur tahun 2012 sebesar 29,01 persen. Kabupaten Probolinggo merupakan kabupaten dengan persentase balita penderita pneumonia yang terendah yaitu sebesar 3,71 persen dan yang tertinggi berada di Kabupaten Bojonegoro sebesar 90,17 persen. Persentase rata-rata balita gizi buruk (X<sub>1</sub>) di Provinsi Jawa Timur sebesar 1,199 persen dengan persentase balita gizi buruk yang tertinggi berada di

Kabupaten Ngawi yaitu sebesar 3,70 persen dan persentase balita gizi buruk yang terendah berada di Kabupaten Jember yaitu sebesar 0,21 persen.

Persentase berat bayi lahir rendah (X<sub>2</sub>) yang tertinggi berada di Kota Madiun yaitu sebesar 9,66 persen. Sedangkan Kabupaten Bangkalan merupakan daerah dengan tingkat persentase berat bayi lahir rendah yang terendah yaitu sebesar 1,25 persen. Kabupaten Malang merupakan kabupaten dengan persentase balita yang mendapatkan imunisasi campak (X<sub>3</sub>) yang tertinggi yaitu sebesar 112,27 persen dan persentase balita yang mendapatkan imunisasi campak yang terendah berada di Kota Kediri yaitu 78,67 persen.

Persentase balita yang mendapatkan vitamin A (X<sub>4</sub>) tertinggi berada di Kota Malang yaitu sebesar 98,87 persen dan Kabupaten Sampang merupakan daerah dengan tingkat persentase balita yang mendapatkan vitamin A terendah yaitu 34,09 persen. Persentase pemberian ASI eksklusif (X<sub>5</sub>) pada balita terendah terdapat di Kabupaten Lamongan yaitu sebesar 37,82 persen dan persentase pemberian ASI eksklusif pada balita tertinggi terdapat di Kabupaten Bangkalan yaitu sebesar 87,08 persen. Kabupaten Sampang merupakan kabupaten dengan persentase balita yang mendapatkan imunisasi DPT+HB (X<sub>6</sub>) tertinggi yaitu sebesar 128,24 persen dan persentase balita yang mendapatkan imunisasi DPT+HB terendah berada di Kota Kediri yaitu 79,86 persen.

#### 3.2. Model Regresi Logistik

Sebelum membentuk suatu model regresi logistik, maka perlu dilakukan uji multikolinieritas untuk mengetahui apakah antar variabel-variabel bebas tidak saling berkorelasi. Untuk mengetahui apakah terjadi multikolinieritas, maka digunakan nilai koefisien korelasi (*Pearson Correlation*). Berdasarkan nilai koefisien korelasi dari masing-masing variabel bebas diperoleh hasil bahwa tidak ada hubungan korelasi yang kuat antar variabel bebas, sehingga variabel-variabel

bebas tersebut dapat digunakan dalam pembentukan model regresi logistik.

Berdasarkan analisis dengan menggunakan regresi logistik, maka model regresi logistik yang terbentuk untuk data risiko penyakit pneumonia pada balita di Provinsi Jawa Timur adalah:

$$g(x_i) = \exp(3,808703 + 0,756180X_2 + 0,447021X_3 - 0,114825X_4 - 0,371864X_6)$$

Berdasarkan model regresi logistik yang terbentuk maka dapat disimpulkan bahwa persentase berat bayi lahir rendah ( $X_2$ ) dan persentase balita yang mendapatkan imunisasi campak ( $X_3$ ) dapat meningkatkan peluang Provinsi Jawa Timur mengalami persentase penderita penyakit pneumonia di atas 27,8 persen. Sedangkan untuk variabel persentase balita yang mendapatkan vitamin A ( $X_4$ ) dan persentase balita yang mendapatkan imunisasi DPT+HB ( $X_6$ ) dapat menurunkan peluang Provinsi Jawa Timur mengalami persentase penderita penyakit pneumonia di atas 27,8 persen.

### 3.3. Model Geographically Weighted Logistic Regression (GWLR)

Langkah pertama dalam mendapatkan model GWLR adalah menentukan letak geografis berdasarkan garis Lintang Selatan dan garis Bujur Timur untuk setiap kabupaten/kota di Provinsi Jawa Timur, menghitung jarak *Euclidean* antara lokasi pengamatan, menentukan nilai *bandwidth* optimum dan menghitung matriks pembobot dari masing-masing titik lokasi. Pembobot spasial yang digunakan pada model GWLR untuk menghitung matriks pembobot tersebut adalah fungsi pembobot *gaussian* kernel. Persamaan dari fungsi pembobot *gaussian* kernel sebagai berikut:

$$w_i(u_i, v_i) = \exp\left(-\frac{1}{2}(d_{ij}/h)^2\right)$$

Pada model GWLR dilakukan pengujian kesesuaian model regresi logistik dengan model GWLR dan pengujian parameter model GWLR secara parsial. Pengujian kesesuaian model

dilakukan untuk mengetahui apakah model GWLR lebih sesuai digunakan dibandingkan dengan model regresi logistik. Hipotesis yang digunakan sebagai berikut:

$H_0 : \beta_k(u_i, v_i) = \beta_k ; k = 1, 2, \dots, k$  (tidak ada perbedaan yang signifikan antara model regresi logistik dan GWLR)

$H_1 : \text{minimal ada satu } \beta_k(u_i, v_i)$  yang berhubungan dengan lokasi  $(u_i, v_i)$  (ada perbedaan yang signifikan antara model regresi logistik dan GWLR)

Tabel 2. Uji Kesesuaian Model Regresi Logistik dan Model GWLR

Model	Devians	df	Devians/df
Regresi Logistik	36,826	31,00	1,188
GWLR	34,371	29,38	1,170

Berdasarkan Tabel 2 didapatkan bahwa nilai  $F_{hitung}$  yang diperoleh dari model regresi logistik dan model GWLR sebesar 1,015384615 apabila digunakan tingkat signifikansi ( $\alpha$ ) sebesar 0,05 maka diperoleh nilai  $F_{tabel}$  sebesar 1,848151713, sehingga dapat disimpulkan bahwa terima  $H_0$  yang artinya tidak ada perbedaan yang signifikan antara model regresi logistik dan GWLR.

Selanjutnya dilakukan pengujian parameter secara parsial. Pengujian dilakukan untuk mengetahui faktor-faktor yang berpengaruh terhadap risiko penyakit pneumonia pada balita di setiap lokasi  $(u_i, v_i)$ . Berdasarkan hasil analisis model GWLR diperoleh hasil bahwa terdapat beberapa variabel yang berpengaruh signifikan di semua titik lokasi yaitu adalah persentase berat bayi lahir rendah ( $X_2$ ), persentase balita yang mendapatkan imunisasi campak ( $X_3$ ), persentase balita yang mendapatkan vitamin A ( $X_4$ ), dan persentase balita yang mendapatkan imunisasi DPT+HB ( $X_6$ ). Selain itu juga terlihat bahwa pada model regresi logistik dan GWLR menunjukkan bahwa pada data risiko penyakit pneumonia pada balita di Provinsi Jawa Timur mengandung kasus overdispersi karena nilai devians dari model regresi logistik dan GWLR yang masing-masing nilai deviansnya dibagi

dengan derajat bebasnya menghasilkan nilai yang lebih besar dari 1.

#### 4. KESIMPULAN DAN SARAN

Berdasarkan hasil analisis dan pembahasan yang dilakukan maka diperoleh kesimpulan tidak ada perbedaan yang signifikan antara model regresi logistik dan GWLR, sehingga dapat disimpulkan bahwa faktor spasial/lokasi tidak berpengaruh terhadap data risiko penyakit pneumonia pada balita di Provinsi Jawa Timur tahun 2012. Model yang dihasilkan regresi logistik dan GWLR menunjukkan bahwa pada data risiko penyakit pneumonia pada balita di Provinsi Jawa Timur tahun 2012 mengandung kasus overdispersi dan variabel-variabel yang berpengaruh signifikan terhadap data risiko penyakit pneumonia pada balita di Provinsi Jawa Timur tahun 2012 adalah persentase berat bayi lahir rendah ( $X_2$ ), persentase balita yang mendapatkan imunisasi campak ( $X_3$ ), persentase balita yang mendapatkan vitamin A ( $X_4$ ), dan persentase balita yang mendapatkan imunisasi DPT+HB ( $X_6$ ).

Adapun saran yang diberikan yaitu untuk mendapatkan model terbaik terhadap risiko penyakit pneumonia pada balita di Provinsi Jawa Timur adalah dengan menggunakan model regresi yang dapat mengatasi adanya kasus overdispersi, misalnya dengan regresi Beta-Binomial dan regresi Logistik Normal, serta pada penelitian berikutnya bisa ditambahkan faktor-faktor dari aspek lain seperti aspek ekstrinsik dan unit pengamatan yang digunakan dapat dipersempit menjadi kecamatan atau kelurahan.

#### DAFTAR PUSTAKA

- [1] Agresti, A. 2002. *Categorical Data Analysis*. Second Edition. New York: John Willy and Sons.
- [2] Anggarini, R. & Puhadi. 2012. Pemodelan Faktor-Faktor Yang Berpengaruh Terhadap Prevalensi Balita Kurang Gizi Di Provinsi Jawa Timur dengan Pendekatan Geographically Weighted Logistic Regression (GWLR). *Jurnal Sains dan Seni ITS*, vol 1, No 1. Surabaya: Institut Teknologi Sepuluh November.
- [3] Fotheringham, A.S., Brunson, C. and Charlton, M. 2002. *Geographically Weighted Regression The Analysis of Spatially Varying Relationships*. University of Newcastle, UK.
- [4] Hosmer, D.W. and Lemeshow, S. 2000. *Applied Logistic Regression*, Second Edition. New York: John Wiley and Sons, Inc.
- [5] Noviana, I., Wulandari, S.P. and Puhadi. 2013. Pemodelan Risiko Penyakit Pneumonia pada Balita di Jawa Timur Menggunakan Regresi Logistik Biner Stratifikasi. *Jurnal Sains dan Seni Pomits*, vol 2.
- [6] Pradita, N.P. 2011. Pemodelan Indeks Pembangunan Manusia (IPM) di Provinsi Jawa Timur dengan Pendekatan Geographically Weighted Logistic Regression. *Skripsi*. Surabaya: Institut Teknologi Sepuluh November. Tidak diterbitkan.
- [7] Santoso, F.P., W, S.P. and Puhadi. 2012. Faktor-Faktor Eksternal Pneumonia Pada Balita di Jawa Timur dengan Peningkatan Geographically Weighted Regression. *Jurnal Sains dan Seni ITS*, vol 1. Surabaya: Institut Teknologi Sepuluh November.